Journal of Medical Virology 51:273–279 (1997)

Molecular Epidemiology of South African Strains of Hepatitis A Virus: 1982–1996

Maureen B. Taylor*

Department of Medical Virology, University of Pretoria, Pretoria, South Africa

Isolates of hepatitis A virus (HAV) are of a single serotype, with human isolates being categorised within four genotypes. In addition, there are three genotypes exclusively associated with Old World monkeys. In some geographical regions, related isolates cluster suggesting endemic spread of the virus, while in other regions several genotypes circulate. Virtually no data are available with regard to the genetic relatedness of South African (SA) strains of HAV. A 177 base segment within the VP1 region and a 168 base segment encompassing the putative VP1/P2A junction of 20 clinical and one environmental wild-type isolate(s) of HAV from SA were amplified by reverse transcriptase-polymerase chain reaction. The nucleotide sequences from the SA isolates showed >85% nucleic acid sequence identity with published sequences for HAV strains from genotype I, with the majority of strains (81%) clustering within subgenotype IB and the remainder in subgenotype IA. A high degree of conservation was noted between the predicted amino acid sequences from SA clinical isolates and isolates from the rest of the world. Data presented indicate that in SA there is a circulating population of endemic HAVs from two distinct subgenotypes. This study provides valuable new data on the genetic relatedness of HAVs from southern Africa and the distribution of subgenotype IB. ***J. Med. Virol. 51:273–279, 1997.***

KEY WORDS: reverse transcriptase-polymerase chain reaction; sequence; genotypes

INTRODUCTION

Due to its unique structural composition, stability characteristics, and tissue tropism, hepatitis A virus (HAV), a hepatotropic virus from the picornavirus family, has been classified and designated type species of the newly created genus *Hepatovirus* [Minor, 1991]. The viral capsid is composed of three exposed polypeptides VP1, VP2, VP3 and a putative VP4, with a highly conserved antigenic structure [Robertson et al., 1992]. Immunological data have shown that there is only one serotype [Rakela et al., 1976; Lemon and Binn, 1983; MacGregor et al., 1983; Dawson et al., 1984], with infection conferring lifelong immunity [Kurstak, 1993]. This has precluded the possibility of using serological approaches to differentiate the sources of HAV in both endemic and epidemic settings. The nucleic acid is a 7.5 kb positive strand RNA molecule which has been shown to be relatively stable [Weitz and Siegl, 1985; Robertson et al., 1992]. However, early studies using two-dimensional mapping of RNase T_1 digests and comparison of sequences indicated that there was significant nucleic acid sequence heterogeneity and that isolates could be grouped according to geographical sources [Weitz and Siegl, 1985]. The cloning and sequencing of the nucleic acid from a number of strains of HAV, including the type strain HM175 [Cohen et al., 1987; Paul et al., 1987; Graff et al., 1994], have facilitated the formulation of primers for the amplification of selected areas of the genome by the reverse-transcriptase polymerase chain reaction (RT-PCR). The amplification and sequencing of variable regions within the capsid proteins, i.e., the C terminus of VP3 through the N terminus of VP1 (VP3/VP1) and the putative VP1/P2A junction of wild-type HAV isolates from different regions of the world, revealed that there was significant nucleic acid sequence heterogeneity, but with limited amino acid heterogeneity [Cohen, 1989; Robertson et al., 1991, 1992; Khanna et al., 1992]. On the basis of their nucleic acid sequence heterogeneity, HAV isolates could be differentiated into seven unique genotypes. A genotype is defined as a group of viruses that differ at no more than 15% of base positions, and a subgenotype as a group of viruses with >92.5% nucleotide sequence identity [Robertson et al., 1992]. Human isolates of HAV can be categorised within four genotypes while three genotypes exclusively are asso-

*Correspondence to: Dr. M.B. Taylor, Department of Medical Virology, Institute of Pathology, P.O. Box 2034, Pretoria, 0001 South Africa.

Accepted 24 September 1996

ciated with Old World monkeys [Robertson et al., 1992]. In some geographical regions, e.g., the United States, Japan, and China, related isolates clustered suggesting endemic spread. In other regions, e.g., Western Europe, HAV isolates were derived from multiple and possibly imported genotypes [Robertson et al., 1992]. Genetic analysis of strains can therefore provide valuable information with regard to the source of the virus in both sporadic and epidemic infection [Kedda et al., 1995; Normann et al., 1995; Apaire-Marchais et al., 1995].

HAV is hyperendemic in South Africa (SA) [Robertson et al., 1992], with epidemiological features of both the developing and the developed world being evident [Martin, 1992]. In the lower socioeconomic and predominantly black African communities, the infection is acquired subclinically in early childhood with about 100% seropositivity by school-going age [Wilkinson, 1983; Prozesky et al., 1984; Abdool Karim and Coutsoudis, 1993; Sathar et al., 1994]. In the higher socioeconomic and predominantly white community, more clinical infections are noted and immunity rises to about 50–70% in adults [Wilkinson, 1983; Prozesky et al., 1984; Prozesky, 1986]. Although a recent investigation reported the presence of subgenotype IA and genotype II amongst SA HAV isolates from selected patients [Kedda et al., 1995], there are very few data on the genotype or genotypes of HAV that are or have been circulating in the SA community [Robertson et al., 1992]. The aim of this investigation was therefore to undertake limited nucleotide sequencing on wild-type strains of HAV from SA to establish their genetic relatedness.

MATERIALS AND METHODS

Specimens

A total of 72 specimens from patients from the former Transvaal province of SA were included in the study. Amongst the specimens were 19 stool specimens from children taken during an outbreak of hepatitis A in a care and rehabilitation centre, 3 stool specimens from cases of sporadic hepatitis A amongst military recruits, and 11 stool specimens taken from children during an outbreak in a crèche. Thirty-three HAV IgM antibody positive sera from both sporadic and epidemic cases of hepatitis A were selected from a bank of sera from patients who had presented with clinical signs or symptoms of hepatitis. Both stool and serum specimens were stored at −20°C with limited freeze-thawing. A sludge sample, taken at a sewerage works servicing predominantly middle to higher socioeconomic communities in the Pretoria area, was included as a source of HAV representing virus currently circulating in the environment.

RNA Extraction

Viral RNA was extracted from 300 μl of a 10% stool suspension in water using the polyethylene glycol-cetyltrimethylammonium bromide method [Jiang et al., 1992]. The extracted RNA was resuspended in a final volume of 20 μl ultrapure water. The sludge sample was treated with chloroform prior to viral recovery by a PEG-NaCl technique [Minor, 1985], followed by RNA extraction as described for the stool samples. RNA was extracted from the HAV IgM positive serum specimens using QIAamp spin columns (QIAamp HCV Kit, Hilden, Germany). A suspension of the cytopathic cell culture adapted strain of HAV, pHM-175 [Cromeans et al., 1987], and ultrapure water, as positive and negative controls, respectively, were included in all procedures.

Oligonucleotide Primers

Two published primer pairs, one amplifying the VP3/VP1 capsid region [Robertson et al., 1991] and the other the putative VP1/P2A junction [Robertson et al., 1992], with expected product sizes of 247 and 360 base pairs, respectively, were used. All primers were synthesised and purified by The Midland Certified Reagent Co. (Midland, TX).

Amplification by RT-PCR

To exclude the possibility of cross-contamination, reagents for the RT-PCR were prepared in a laminar flow cabinet. The RT-PCR reactions were carried out in rooms separate from those used for the initial processing of the specimens and analysis of the amplicons.

Reverse transcription was performed in a 49 μl reaction containing 5 μl extracted RNA, 5 units of avian myeloblastosis virus RT (Promega Corp., Madison, WI), 1 × PCR buffer (10 mM Tris-HCl [pH 8.8], 25 mM KCl, 3.5 mM $MgCl_2$), 100 ng reverse primer, and 40 units ribonuclease inhibitor (rRNasin: Promega Corp.). The reaction mix was incubated at 42°C for 1 hr. After incubation, 100 ng of the forward primer and 1.5 units *Taq* DNA polymerase (Boehringer Mannheim GmbH) were added. The steps of the PCR cycling program included denaturation at 95°C for 1 min, annealing at 45°C for 1 min, and extension at 72°C for 75 sec for 30 cycles. PCR products were analysed by polyacrylamide gel electrophoresis and visualised by ethidium bromide staining and UV illumination.

Sequencing of RT-PCR Amplified Products

RT-PCR products, i.e., amplicons from the VP3/VP1 region of 21 SA isolates and from the VP1/P2A region of 14 of the aforementioned 21 SA isolates and the positive control, were sequenced directly using the Sequenase PCR Product Sequencing Kit (U.S. Biochemical Corp., Cleveland, OH). The sequencing reactions were run on an 8% polyacrylamide-6 M urea gel in 1 × TBE buffer. Gels were vacuum dried and exposed to X-ray film for 12 hr at room temperature.

Sequence and Alignment Data Analysis

The sequences of the cDNA HAV amplicons were compared to published sequences for HAV strains [Robertson, personal communication; Robertson et al., 1987, 1991, 1992; Cohen et al., 1987; Paul et al., 1987; Graff et al., 1994; Apaire-Marchais et al., 1995], using

TABLE I. Demographic Patient and Source Data of the SA HAV Isolates

Year	Isolate	Specimen	Patient data		Epidemiological data
			Age	Ethnicity[a]	
1982	VDM[b,c]	Stool	Adult	W	Military recruit
1983	JVR[b,c]	Stool	Adult	W	Military recruit
1984	2334[b]	Serum	Adolescent	W	} Outbreak: home for the
	2328[b,c]	Serum	Adolescent	W	} mentally retarded
	2333[b,c]	Serum	Adolescent	W	}
1989	923359[b,c]	Stool	7 years	W	} Outbreak: care and
	923200[b,c]	Stool	7 years	W	} rehabilitation centre
1992	6100[b]	Stool	2 years	W	} Outbreak: Crèche
1993	314274[b,c]	Serum	2,5 years	A	Sporadic
1994	406443[b]	Serum	14 years	A	Sporadic
	406808[b,c]	Serum	44 years	W	Sporadic
	406909[b,c]	Serum	6,5 years	A	Sporadic
	412991[b,c]	Serum	6 years	A	Sporadic
	415643[b]	Serum	2 years	A	Sporadic
1995	503712[b,c]	Serum	18 years	W	Sporadic
	504184[b,c]	Serum	20 years	W	Sporadic
1996	96001190[b,c]	Serum	12 years	W	} Outbreak: Family
	96002041[b]	Serum	11 years	W	}
	96002382[b,c]	Serum	28 years	W	Sporadic
	96002054[b]	Serum	6 years	A	Sporadic
	SLUDG1/96[b]	Sludge			Daspoort Sewerage Works

[a]W, European origin (white Caucasian) and predominantly higher socioeconomic group; A, African ethnic origin and predominantly lower socioeconomic group.
[b]Isolates included in phylogenetic analysis based on N-terminal VP1 capsid protein region.
[c]Isolates included in phylogenetic analysis based on putative VP1/P2A junction.

the CLUSTAL program of the PC/GENE software, version 6.8 (IntelliGenetics, Mountain View, CA).

Nucleotide Sequence Accession Numbers

The nucleotide sequence data of the SA HAV isolates have been deposited with the EMBL/GenBank database and assigned the following accession numbers: 412991 (U68689), 96001190 (U66482; U68690), 406909 (U66483; U68691), JVR (U68692), 503712 (U68693), 9602382 (U68694), 406808 (U68695), 923359 (U68696; U66481), VDM (U66489; U68697), 2333 (U66487; U68698), 314274 (U66484; U68699), 504184 (U66485; U68700), 415643 (U66486), and SLUDG1/96 (U66488).

RESULTS

Detection and Analysis of the SA Wild-Type HAV Isolates

RT-PCR yielded products of the expected size from 6 of the stool specimens and from 15 of the serum specimens tested with both primer pairs. HAV RNA, but only from the VP3/VP1 capsid region, was also amplified directly from the sludge sample. Multiple bands were not observed in any of the lanes.

Nucleic Acid Sequence Analysis of the SA Wild-Type HAV Isolates

The chronological and demographic patient data and sources of the SA HAV isolates characterised are shown in Table I. Pairwise analysis of the cDNA nucleic acid sequences of a 177 base region at the N terminus of the VP1 capsid protein of 21 of the SA isolates with one another and the equivalent regions of HAVs representative of all genotypes and subgenotypes revealed that the SA isolates characterised all showed >85%, i.e., 89.8–96.1% identity with the sequence of the type strain, HM175 (Fig. 1). Isolates from the same outbreak were all found to be identical and consequently only a representative strain is included. The cytopathic variant of HM175, pHM175, showed 100% nucleotide sequence identity with the parent strain in this region (data not shown). Isolates from an outbreak of hepatitis A in adolescents in a home for the mentally retarded in 1984, i.e., 2334, 2333, and 2328, were found to be identical to the isolate JVR from a military recruit in 1983. These isolates in turn were found to show 99% nucleotide sequence identity with isolates 503712 and 96002382 from sporadic cases of hepatitis A in adults in 1995 and 1996, respectively. Isolates from a family outbreak in a middle to higher socioeconomic community in 1996, i.e., 96001190 and 96002041, showed 100% nucleotide sequence identity with an isolate, 412991, from a child from a lower socioeconomic community who presented with hepatitis in 1994. The latter isolates in turn showed 99% nucleotide sequence identity with isolates from a crèche outbreak in 1992 (6100) and a sporadic case of hepatitis A in 1993 (314274). SA isolates 96001190, 96002041, 6100, 314274, 406909, 412991, and 415643 showed a higher percentage nucleotide sequence identity (>96–98%) to isolate Jor88 from Amman, Jordan, than to the type strain HM175 from Australia, or isolate MBB from North Africa (94–96%). Isolate 406808, from a case of sporadic hepatitis A in an adult in 1994, showed 99.4% nucleotide sequence identity with isolate GA88 from a sporadic case of hepatitis A in Atlanta, Georgia, in 1988. Identical isolates 923300 and 923359 showed a higher percentage nucleotide sequence identity (⩾96%) with isolates from China (China81, China82, China83,

```
          2208                                                  2257
HM175     GTTGGAGATG  ATTCTGGAGG  TTTTTCAACA  ACAGTTTCTA  CAGAACAGAA
96001190  ..........  ..........  ..........  ..........  ....g.....
412991    ..........  ..........  ..........  ..........  ....g.....
6100      ..........  ..........  ..........  ..........  ....g.....
96002054  ..........  ..........  ..........  ..........  ..........
314274    ..........  ..........  ..........  ..........  ....g.....
406909    ..........  ..........  ..........  ..........  ....g.....
504184    ..........  ..........  ..........  ..........  ....g.gt..
406443    ..........  ..........  ..........  ..........  ....g.....
415643    ..........  ..........  ..........  ..........  ....g.....
JVR       ..........  ..........  ...c......  ........c.  ....g.....
2333      ..........  ..........  ...c......  ........c.  ....g.....
96002382  ..........  ..........  ...c......  ........c.  ....g.....
503712    ..........  ..........  ...c......  ........c.  ....gt....
SLUDG1/96 ..........  ..........  ...c......  ..........  ....g.....
406808    ..........  ....a.....  ...c......  ..........  ....g.....
923359    ..........  ....a..g..  ..........  ........c.  ....g.....
VDM       ..........  ....a.....  ..........  ........c.  ....g.....
Consensus **********  ****-**-**  ***-******  *********-  ****----**

          2258                                                  2307
HM175     TGTTCCAGAT  CCCCAAGTTG  GTATAACAAC  CATGAAAGAT  TTGAAAGGAA
96001190  ..........  ..t.......  ..........  ......g...  ..a.......
412991    ..........  ..t.......  ..........  ......g...  ..a.......
6100      ..........  ..t.......  ..........  ......g...  ..a.....t.
96002054  ..........  ..t.......  ..........  ......g...  ..a.....t.
314274    .a........  ..t.......  ..........  ......g...  ..a.....t.
406909    ..........  ..t.......  ..........  ......g...  ..a.....t.
504184    ..........  .....g....  ..........  ......g...  ..a.......
406443    ..........  .....g....  ..........  ......g...  ..a.......
415643    ..........  ..........  .......g..  .g....g...  ..ag.g....
JVR       ..........  ..t.......  ..........  ......g...  ..a.....g.
2333      ..........  ..t.......  ..........  ......g...  ..a.....g.
96002382  ..........  ..t.......  ..........  ......g...  ..a.....g.
503712    ..........  ..t.......  ..........  ......g...  ..a.....g.
SLUDG1/96 ..........  ..t.......  ..........  ......g...  ..a.......
406808    ......t...  ..........  ..........  ......g...  ..a.....g.
923359    ......t...  ........c.  .c........  .....gg..c  c.a.....g.
VDM       ......t...  .....g..c.  .c........  .....gg...  c.a.....g.
Consensus *-****-***  **-**-**-*  *-******-**  *-***--**-  -*--*-**-*

          2308                                                  2357
HM175     AAGCTAACAG  AGGGAAAATG  GATGTTTCAG  GAGTACAAGC  ACCTGTGGGA
96001190  ....c..t..  ..........  ..........  ....g.....  ..........
412991    ....c..t..  ..........  ..........  ....g.....  ..........
6100      ....c..t..  ..........  ..........  ....g.....  ..........
96002054  ....c..t..  ..........  ..........  ....g.....  ..........
314274    ....c..t..  ..........  ..........  ....g.....  ..........
406909    .......t..  ..........  ..........  ....g.....  ..........
504184    ....c.....  ...c......  ..........  ....g.....  ..........
406443    ....c..t..  ...c......  ..........  ....g.....  ..........
415643    ....c..t..  ..........  ..........  ....g.....  ..........
JVR       .g........  ...a......  ..........  ....g.....  ..........
2333      .g........  ...a......  ..........  ....g.....  ..........
96002382  .g........  ...a......  ..........  ....g.....  ..........
503712    .g........  ...a......  ..........  ....g.....  ..........
SLUDG1/96 cg.....t..  ...a.c....  ..........  ....g.....  ..........
406808    ....c..t..  g..a..g...  .....a....  ....g.....  ..........
923359    ....c..t..  ...a......  ..........  ....c.....  ..........
VDM       .......t..  g..a..g...  ..........  ....g.....  ..........
Consensus *-**-**-**  -**-*--***  *****-****  ****-*****  **********

          2358          2384
HM175     GCTATCACAA  CAATTGAGGA  TCCAGTT
96001190  ..........  ..........  ......c      95.5%
412991    ..........  ..........  ......c      95.5%
6100      ..........  ..........  ......c      94.9%
96002054  ..........  ..........  .......      96.1%
314274    ..........  ..........  ......c      94.4%
406909    .....g....  ..........  ......c      94.4%
504184    ..........  ..........  .......      94.9%
406443    ..........  ..........  .......      95.5%
415643    ..........  ..........  .......      94.4%
JVR       ..........  ..........  ......c      93.8%
2333      ..........  ..........  ......c      93.8%
96002382  .....t....  ..........  ......c      93.2%
503712    ..........  ..........  ......c      93.2%
SLUDG1/96 ..........  ......c...  ......c      92.7%
406808    .....t....  ..........  ......g      91.0%
923359    ..........  .......a..  .......      89.8%
VDM       ..........  ..........  .......      90.4%
Consensus ******-****  *******--**  *******-
```

Fig. 1. cDNA nucleotide sequence obtained from the HAV amplicons of the first 177 bases of the N-terminal VP1 capsid protein region of SA wild-type isolates of HAV compared to the published sequence for the type strain, HM175. Nucleotides identical to HM175 are shown by dots; variable nucleotides are shown in lowercase letters; Asterisks denote consensus among all isolates. Percentage identity to HM175 is shown at the end of the sequence.

China88) and the former USSR (RUS1406) than to any other isolates investigated (results not shown). Isolate VDM showed 96.9% nucleotide sequence identity with a cell culture passaged variant of MS-1, a virus collected over 30 years ago and used in the Willowbrook human volunteer studies [Krugman et al., 1962]. The environmental isolate, SLUDG1/96, showed the highest percentage nucleotide sequence identity (96.6%) to SA isolates JVR and 2333/2334/2328 collected in 1983 and 1984, respectively, and 96% identity to isolates 503712 (1995) and 96002382 (1996).

Within the 168 base region encoding the putative VP1/P2A junction, the nucleotide sequences of isolates 412991 and 96001190 were again found to be identical. These isolates also showed a higher percentage nucleotide sequence identity to HM175 in this region than in the N terminus of the VP1 capsid protein region, i.e., 98.2% compared to 95.5%. SA isolates 314274 and 406909 also showed a higher percentage nucleotide sequence identity to HM175 in this region, i.e., 97.6% and 96.4%, respectively, than in the N terminus of the VP1 capsid protein region. As was demonstrated for the N terminus of the VP1 capsid protein region, SA isolates JVR, 2333, and 2328 and isolates 923359 and 923300 were found to show 100% nucleotide sequence identity in the putative VP1/P2A junction. However, their percentage nucleotide sequence identities to HM175 were not markedly different to those recorded for the N-terminal VP1 capsid protein region.

Epidemiological Analysis of the SA Wild-Type Isolates

The genetic relatedness of the SA isolates, based on the 177 base region at the N terminus of the VP1 capsid protein, is shown in the dendrogram (Fig. 2). Only a representative strain of identical isolates from the same outbreak is included. The SA isolates characterised all clearly cluster within genotype I. The isolates from the outbreak in the care and rehabilitation centre in 1989 (923359 and 923300), from one of the military recruits (VDM) in 1982, and a sporadic case (406808) in 1994 all group within subgenotype IA, but are clearly distinct from one another. The remaining SA isolates all cluster within subgenotype IB. Several of the latter strains grouped in two distinct clusters, but were dispersed with regard to year of derivation and relatedness to strains HM175 and Jor88.

The genetic relatedness, based on a 168 base region at the putative VP1/P2A junction, of 14 of the 21 SA isolates characterised for this region showed identical genogroupings to those shown for the N terminus of the VP1 capsid protein (results not shown).

Amino Acid Sequence Analysis of the SA Wild-Type Isolates

Pairwise analysis of the 59 amino acids at the N terminus of the VP1 capsid protein of the 20 clinical isolates characterised showed 97–100% amino acid sequence identity with HM175, with the majority (70%) of the clinical isolates being identical to HM175. The

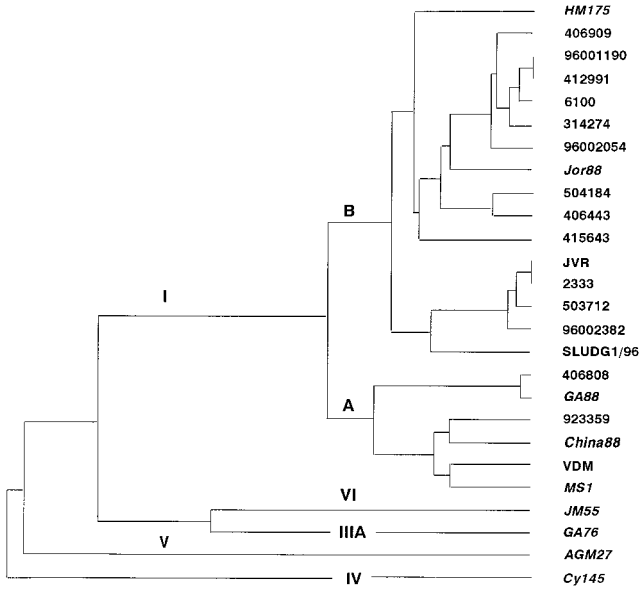

Fig. 2. Dendrogram of a pairwise comparison of the nucleotide sequences (nucleotides 2208–2384) of the N-terminal VP1 capsid protein region of HAV isolates of genotypes and subgenotypes IA (GA88, China88, MS1), IB (HM175, Jor88), IIIA (GA76), IV (Cy145), V (AGM-27), VI (JM55), and the wild-type SA HAV isolates. Roman numerals designate the genotypes, and A and B designate subgenotypes.

environmental isolate (SLUDG1/96), however, showed only 95% amino acid sequence identity with HM175. The differences in the predicted amino acid sequences of the seven varying, isolates, compared to that of HM175, are shown in Figure 3a. All isolates show a single amino acid substitution except for isolates 415643 and SLUDG1/96, where two and three substitutions were noted, respectively. As is evident from Figure 3a, no pattern was noted in the substitutions except in the isolates from subgenotype IA, where the substitution was noted in the same position, i.e., Lys 29 of VP1.

Pairwise analysis of the 56 amino acids at the putative VP1/P2A junction of 14 of the SA clinical isolates showed >98–100% amino acid sequence identity with HM175. The differences in the predicted amino acid sequences of the five varying isolates, compared to that of HM175, are shown in Figure 3b. Compared to HM175, four isolates show a single amino acid substitution, i.e., at Met 277 of VP1, with isolates from subgenotype IA showing 100% identity with the published consensus sequence for subgenotype 1A [Robertson et al., 1992]. Isolate 504184 also showed a single amino acid substitution, i.e., at Cys 2A-5, similar to that seen for isolates CF53/Berne and SLF88 (genotypes II and VII, respectively) [Robertson et al., 1992].

DISCUSSION

In this investigation, RT-PCR and sequence analysis were used to determine the genetic relatedness of SA strains of HAV. In order to achieve any significant epidemiological groupings, it is essential to use wild-type virus to circumvent any changes associated with cell culture or animal adaptation [Robertson et al., 1992]. Consequently, HAV nucleic acid was amplified and sequenced directly from original SA specimens rather than from cell culture or animal-passaged material. The results indicate that HAVs associated with human infection in SA during the period investigated clustered within genotype I. Preliminary results suggested that subgenotype IA predominated with only a single isolate from subgenotype IB being identified [Taylor and Wolfaardt, 1995]. This was not unexpected as genotype I comprises 80% of the human HAV strains studied, with subgenotype IA being the most prominent subgenotype identified [Robertson et al., 1992]. Subgenotype IA constitutes the major virus population in both North and South America, China, Japan, the former USSR, and Thailand, with additional strains being more randomly dispersed with respect to geographical origin [Robertson et al., 1992]. However, as more SA isolates were characterised, it became evident that subgenotype IB was more prominent, constituting 81% of the isolates characterised in this study. Subgenotype IB contains strains from Jordan, North Africa, Australia, Europe, Japan, and South America, with the majority of strains recovered from areas near the Mediterranean [Robertson et al., 1992]. These data suggest that there is a circulating population of endemic HAVs from two distinct subgenotypes in this region of SA, with subgenotype IA being less prominent. This cocirculation of subgenotypes IA and IB in this region is similar to what has recently been reported for western France [Apaire-Marchais et al., 1995].

A high degree of genomic conservation between SA isolates and isolates from the rest of the world has been demonstrated, confirming the genetic stability of HAV as noted in previous studies [Robertson et al., 1992], in comparison to that of poliovirus [Kew et al., 1984]. Identical isolates were also identified in specimens collected from different areas of the Transvaal in 1983 (JVR) and 1984 (2333, 2334, 2328), and in 1994 (412991) and 1996 (96001190 & 96002041), with a number of closely related isolates collected from other areas of the Transvaal at differing times clustering together with these identical isolates (Fig. 2). To ensure that these results were not due to PCR contamination, additional extractions, separate RT-PCR amplification, and sequencing reactions were carried out. The occurrence of HAV strains with >99% nucleotide sequence identity from different geographical regions and continents, as found with SA isolates 406808 and GA88, has been noted previously [Robertson et al., 1992].

The SA isolates do not differ greatly in their predicted amino acid sequence to that of the type strain, HM175. Strain HM175 is used as the source virus for the vaccine strain for the HAV vaccine currently licensed for use in SA. Vaccine-induced immunity should therefore be protective against HAVs currently circulating in SA.

In their investigation of SA isolates of HAV, Kedda et al. [1995] reported the aetiological association of HAV genotype IA with an outbreak of hepatitis A due to

```
            VP1

            1                            29                           59

HM175       VGDDSGGFSTTVSTEQNVPDPQVGITTMKDLKGKANRGKMDVSGVQAPVGAITTIEDPV
406909      ...................................................M.......      IB
314274      .................I.........................................      IB
504184      ...............R...........................................      IB
415643      ...........................V...E...........................      IB
SLUDG1/96   .................................T....T................A...      IB
923359      ............................R..............................      IA
VDM         ............................R..............................      IA

a Consensus ***************_*_*********__**_*_****_************_***_***

            VP1                         →P-2A
            252                    300          10        20

HM175       ESMMSRIAAGDLESSVDDPRSEEDKRFE  SHIECRKPYKELRLEVGKQRLKYAQEEL
406909      ........................H...  ............................ IB
504184      ............................  ....S....................... IB
406808      ........................R...  ............................ IA
923359      ........................R...  ............................ IA
VDM         ........................R...  ............................ IA
b Consensus ************************_***  ****_***********************
```

Fig. 3. **a:** Comparison of the predicted amino acid sequences at the N-terminal VP1 capsid protein region of the SA wild-type HAV isolates compared to the published sequence for the type strain, HM175. Amino acids identical to HM175 are shown by dots; variable amino acids are shown by the appropriate single letter amino acid code; asterisks denote consensus amongst all isolates. Only a representative strain of identical isolates from the same outbreak is shown. Genetic relatedness is shown at the end of the sequence. **b:** Comparison of the predicted amino acid sequences at the putative VP1/P2A junction of the SA wild-type HAV isolates compared to the published sequence for the type strain, HM175. Amino acids identical to HM175 are shown by dots; variable amino acids are shown by the appropriate single letter amino acid code; asterisks denote consensus amongst all isolates. Only a representative strain of identical isolates from the same outbreak is shown. Genetic relatedness is shown at the end of the sequence.

contaminated factor VIII, and with a sporadic case of HAV infection. They also reported the presence of HAV genotype II in specimens from both sporadic cases and a crèche outbreak of hepatitis A. This is of great significance as only one isolate of genotype II, i.e. CF53/Berne from Clermont-Ferrand, France, appears to be documented in the literature [Jansen et al., 1990; Robertson et al., 1992], and no genotype II isolates were detected in this investigation. The isolates characterised by Kedda et al. [1995], as well as those in this investigation, were predominantly also all from the former Transvaal province. However, no data with regard to the possible source of the infection or travel history were available from patients in this investigation. Therefore, additional isolates from different communities from different geographical regions within SA will have to be characterised to ascertain whether or not similar strains are circulating in those regions. With increasing urbanisation and consequent expected changes in the epidemic vulnerability of the SA population to HAV, clinical HAV infection may become more prevalent [Prozesky, 1986; Martin, 1992]. This study, therefore, provides essential baseline data for genetic analysis of strains during potential outbreaks to determine and identify the source of the virus.

To our knowledge this is the most comprehensive study on the genetic relatedness of HAVs from southern Africa and provides valuable new data on the distribution of subgenotype IB.

ACKNOWLEDGMENTS

Thanks are due to Mrs. Nerine Cox for technical assistance; Mrs. Marianne Wolfaardt for advice and support; Dr. BH Robertson, Hepatitis Branch, CDC, Atlanta, GA, for providing HAV sequences and reviewing the manuscript; and Prof WOK Grabow for affording me the opportunity to do this investigation. This work was supported by grants from the South African Poliomyelitis Research Foundation, the Faculty of Medicine of the University of Pretoria, South Africa, and the Water Research Commission.

REFERENCES

Abdool Karim SS, Coutsoudis A (1993): Sero-epidemiology of hepatitis A in black South African children. South African Medical Journal 83:748–750.

Apaire-Marchais V, Robertson BH, Aubineau-Ferre V, le Roux MG, Leveque F, Schwartzbrod L, Billaudel S (1995): Direct sequencing of hepatitis A virus strains isolated during an epidemic in France. Applied and Environmental Microbiology 61:3977–3980.

Cohen JI (1989): Hepatitis A virus: Insights from molecular biology. Hepatology 9:889–895.

Cohen JI, Ticehurst JR, Purcell RH, Buckler-White A, Baroudy BM (1987): Complete nucleotide sequence of wild-type hepatitis A virus: Comparison with different strains of hepatitis A virus and other picornaviruses. Journal of Virology 61:50–59.

Cromeans T, Sobsey MD, Fields HA (1987): Development of a plaque assay for a rapidly replicating, cytopathic hepatitis A virus. Journal of Medical Virology 45:45–56.

Dawson GJ, Decker RH, Norton DK, Bryce DK, Whittingon RO, Tribby II, Mushahwar IK (1984): Monoclonal antibodies to hepatitis A virus. Journal of Medical Virology 14:1–8.

Graff J, Normann A, Feinstone SM, Flehmig B (1994): Nucleotide sequence of wild-type hepatitis A virus GBM in comparison with two cell culture-adapted variants. Journal of Virology 68:548–554.

Jansen RW, Siegl G, Lemon SM (1990): Molecular epidemiology of human hepatitis A virus defined by an antigen-capture polymerase chain reaction. Proceedings of the National Academy of Sciences USA 87:2867–2871.

Jiang X, Wang J, Graham DY, Estes MK (1992): Detection of Norwalk virus in stool by polymerase chain reaction. Journal of Clinical Microbiology 30:2529–2534.

Kedda M-A, Kew MC, Cohn RJ, Field SP, Schwyzer R, Song E, Fernandes-Costa F (1995): An outbreak of hepatitis A among South African patients with hemophilia: Evidence implicating contaminated factor VIII concentrate as the source. Hepatology 22:1363–1367.

Kew OM, Nottay BK, Obijeski JF (1984): Applications of oligonucleotide fingerprinting to the identification of viruses. In Maramorosch K, Koprowski H (eds): "Methods in Virology," vol 8. New York: Academic Press, pp 41–84.

Khanna B, Spelbring JE, Innis BL, Robertson BH (1992): Characterization of a genetic variant of human hepatitis A virus. Journal of Medical Virology 36:118–124.

Krugman S, Ward R, Giles JP (1962): The natural history of infectious hepatitis. American Journal of Medicine 32:717–728.

Kurstak E (1993): "Viral Hepatitis. Current Status and Issues." Wien: Springer-Verlag, p 44.

Lemon SM, Binn LN (1983): Antigenic relatedness of two strains of hepatitis A virus determined by cross-neutralization. Infection and Immunity 42:418–420.

MacGregor A, Kornitschuk M, Hurrell JGR, Lehmann NI, Coulepis AG, Locarnini SA, Gust ID (1983): Monoclonal antibodies against hepatitis A virus. Journal of Clinical Microbiology 18:1237–1243.

Martin DJ (1992): Hepatitis A vaccination—an option for South Africa? South African Medical Journal 82:5–6.

Minor PD (1985): Growth, assay and purification of Picornaviruses. In Mahy BWJ (ed): "Virology: A Practical Approach." Oxford: IRL Press Ltd., pp 25–41.

Minor PD (1991): Picornaviridae. Classification and nomenclature of viruses: Fifth report of the International Committee on Taxonomy of Viruses. Archives of Virology Supplement 2:320–326.

Normann A, Pfisterer-Hunt M, Schade S, Graff J, Chaves RL, Crovari P, Icardi G, Flehmig B (1995): Molecular epidemiology of an outbreak of hepatitis A in Italy. Journal of Medical Virology 47:467–471.

Paul AV, Tada H, von der Helm K, Wissel T, Kiehn R, Wimmer E, Deinhardt F (1987): The entire nucleotide sequence of the genome of human hepatitis A virus (isolate MBB). Virus Research 8:153–171.

Prozesky OW (1986): Viral hepatitis in southern Africa. South African Medical Journal "Fetschrift Supplement" 70:55–57.

Prozesky OW, Whitcutt JM, Coppin A, Rossouw E (1984): Diagnosis of viral hepatitis. Experience in a South African laboratory. South African Medical Journal 65:591–594.

Rakela J, Fay OH, Stevenson D, Gordon I, Mosley JW (1976): Similarities of two hepatitis A virus strains. Bulletin of the World Health Organization (Geneva) 54:561–564.

Robertson BH, Brown VK, Bradley DW (1987): Nucleic acid sequence of the VP1 region of attenuated MS-1 hepatitis A virus. Virus Research 8:309–316.

Robertson BH, Khanna B, Nainan OV, Margolis HS (1991): Epidemiologic patterns of wild-type hepatitis A virus determined by genetic variation. Journal of Infectious Diseases 163:286–292.

Robertson BH, Jansen RW, Khanna B, Totsuka A, Nainan OV, Siegl G, Widell A, Margolis HS, Isomura S, Ito K, Ishizu T, Moritsugu Y, Lemon SM (1992): Genetic relatedness of hepatitis A virus strains recovered from different geographical regions. Journal of General Virology 73:1365–1377.

Sathar MA, Soni PN, Fernandes-Costa FJTD, Wittenberg DF, Simjee AE (1994): Racial differences in the seroprevalence of hepatitis A virus infection in Natal/Kwazulu, South Africa. Journal of Medical Virology 44:9–12.

Taylor MB, Wolfaardt M (1995): Genetic characterization of South African strains of hepatitis A virus. Johannesburg: South Africa: Abstracts of the Vth International Congress on the Impact of Viral Diseases in the Developing World, p 69.

Weitz M, Siegl G (1985): Variation among hepatitis A strains. I. Genomic variation detected by T_1 oligonucleotide mapping. Virus Research 4:53–67.

Wilkinson R (1983): The frequency of hepatitis A and B viruses as the offending viral type in suspected hepatitis. South African Medical Journal 63:82–85.